Portfolio Methods in Theorem Proving for Elementary Geometry

Vesna Marinković[1] Mladen Nikolić[1] Zoltán Kovács[2]
Predrag Janičić[1]

[1]Faculty of Mathematics, University of Belgrade, Serbia

[2]The Private University College of Education of the Diocese of Linz, Austria

ADG 2016, Strasbourg, France, June 27-29, 2016.

Automated Reasoning in Geometry

Automated reasoning in geometry:

- Automated theorem proving
- Interactive theorem proving
- Solving geometry constraints
- Discovery
- Dynamic geometry systems
- ...

And Now for ....
...Something Completely Different...

Automated reasoning in geometry
with
Machine learning

Overview

Portfolio Problem Solving

- General context:
  - There are several solvers/provers available
  - For each input instance, one solver is to be selected
- Selection process is based on some specifics (values of some features) of the input instance
- Expected benefits:
  - more solved instances than for any individual solver
  - overall solving time lower than for any individual solver
- Successfully used in different areas, especially in SAT solving

Portfolio Problem Solving in Geometry

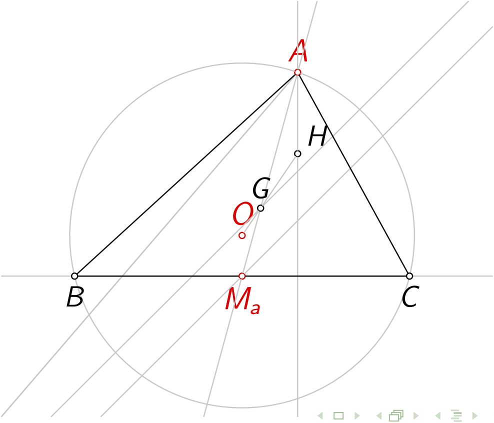
- The portfolio approach has not been used in proving geometry theorems so far
- Useful for systems that can use several provers (GeoGebra, GCLC, ArgoTriCS)
- Challenges:
  - no standard format for storing geometry theorems
  - different provers implemented in different languages
  - different domains of geometry provers
  - identifying suitable features

Corpus of Theorems

- Theorems obtained from solving construction problems:
  - 560 triangle location problems from Wernick's list
  - Example: given A, O and M_a construct the triangle ABC
- It has to be proved that points given by the problem setting are the corrsponding points of the constructed triangle
- For each solution, there are three correctness theorems
- Solutions to construction problems and correctness theorems generated automatically by our tool ArgoTriCS

Running Example (illustration of the construction)

Problem 28: Given a point A, a point O and a point M_a construct the triangle ABC.



Running Example (correctness theorem)

Construction:

1. Using the point A and the point M_a, construct a point G (rule W01);
2. Using the point O and the point G, construct a point H (rule W01);
3. Using the point A and the point H, construct a line h_a (rule W02);
4. Using the point A and the point O, construct a circle $k(O, C)$ (rule W06);
5. Using the point M_a and the line h_a, construct a line a (rule W10);
6. Using the circle $k(O, C)$ and the line a, construct a point C and a point B (rule W04);
7. Using the point B and the point C, construct a point $\_M_a$ (rule W01);
8. Using the point A and the point C construct the line $\_b$ (rule W02);
9. Using the point C and the point A, construct a point $\_M_b$ (rule W01);
10. Using the point B and the point C construct the line $\_a$ (rule W02);
11. Using the point $\_M_a$ and the line $\_a$ construct the line $\_m_a$ (rule W10);
12. Using the point $\_M_b$ and the line $\_b$ construct the line $\_m_b$ (rule W10);
13. Using the line $\_m_a$ and the line $\_m_b$ construct the point $\_O$ (rule W03);

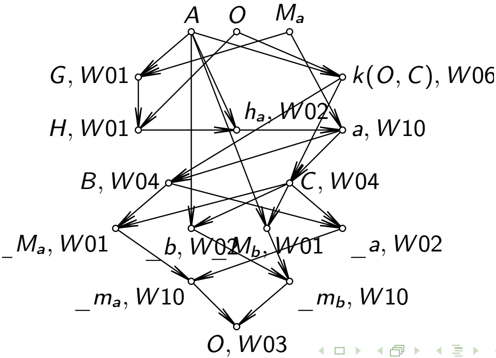
Statement: Prove that the point M_a is identical to the point $\_M_a$.

Results for Individual Provers

- For each theorem four theorem provers used:
  - GCLC (Area method, simple Wu's method, GB method)
  - OpenGeoProver (simple Wu's method)
- Four possible outcomes:
  1. Conjecture proved within time limit (5 minutes)
  2. Conjecture is out of the scope of a prover
  3. The prover can neither prove nor disprove the conjecture
  4. Timeout reached without getting an answer
- Focus on problems that ArgoTriCS succeeded to solve:
  - 828 conjectures in total
  - 537 proved by at least one prover
  - Best prover (OGP-Wu) proved 486 conjectures
  - Room for portfolio: 537-486=51

Construction Graph

- A node for each given or constructed object
- Two nodes are connected by edge if the object from the first node is used in the construction step for the second node



Set of Features over Construction Graph

Basic features:

- the number of nodes/edges in the graph
- the ratio of number of nodes and edges and its reciprocal
- the ratio of longest path length and number of nodes/edges and its reciprocal
- the node degree statistics
- the rule application frequencies (normalized) and their statistics
- the object type frequencies (normalized) and their statistics ...

Additional features (statistics over the basic features):

- the mean, variation coefficient, minimum, maximum, entropy of distributions

Two Portfolio Systems

1. Portfolio system based on the k nearest neighbours technique
   - the training set consists of feature vectors and proving times for each prover
   - for the input instance, k nearest ones from the training set are found
   - the prover with best performance on these k instances is selected
2. Portfolio system based on multinomial logistic regression
   - in the training phase, prover desirability for each instance is based on their relative proving times for that instance
   - the prover with highest probability predicted is selected

Conditions under which Evaluation is Performed

- Proving and exploiting cutoff time is set to 5 minutes
- Evaluation is performed by 5-fold nested crossvalidation
- Since the set of problems is randomly split into subsets, results are averaged over 100 runs
- Reference provers:
  - the best individual prover among available provers
  - the virtual best prover

Results

Prover	# proved	Time (s)
Best prover	486.0	15831
MLR portfolio	525.4	4250
k-nn portfolio	526.5	3776
Virtual best prover	537.0	776

- Two portfolios perform almost the same

Predicting whether prover will finish in cutoff time

Predicting whether prover will finish within the cutoff time:

- The problem modeled as a classification problem (yes/no output)
- Regularized logistic regression used for classification
- Products of features are added to the feature set to model feature interactions
- The classification accurace ranges from 94% to 98% depending on the prover

Predicting prover runtime

Predicting prover runtime:

- Runtimes range from 0s to 184s
- The ridge regression used for prediction
- Nested crossvalidation used for evaluation
- For three provers, root mean square error ranges from 0.07s to 3.07s
- For prediction of logarithm, root mean square error ranges from 0.04s to 0.39s

Current and Future Work

- Portfolio approach in geometry theorem proving performs well
- Currently we develop similar protfolio for provers built in GeoGebra
- We plan to consider other corpora of geometry theorems (e.g., Connelly's corpus)
- We plan to investigate relationship between feature values and prover to be selected